

**UNITED STATES DISTRICT COURT  
SOUTHERN DISTRICT OF NEW YORK**

-----	X	
Monique Da Silva Moore, et al.,	:	
	:	
Plaintiffs,	:	
	:	
v.	:	11 Civ. 1279 (ALC)(AJP)
	:	
Publicis Groupe SA and MSL Group,	:	
	:	
Defendants	:	
-----	X	

**DECLARATION OF JAN PUZICHA IN SUPPORT OF DEFENDANT  
MSLGROUP'S OPPOSITION TO PLAINTIFFS' RULE 72(A) OBJECTION  
TO MAGISTRATE JUDGE PECK'S FEBRUARY 8, 2012 RULINGS**

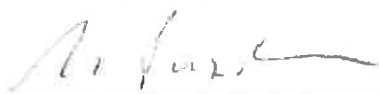
Jan Puzicha, being duly sworn, states as follows:

1. I am a resident of Germany.
2. I am the Chief Technology Officer at Recommind, Inc. Recommind has been engaged by Defendant MSL Group ("MSL") to provide eDiscovery technology to assist MSL and its legal counsel.
3. I hold a Ph.D. from the University of Bonn in Computer Science, from which I graduated with honors. I am an inventor on the patent for Probabilistic Latent Semantic Analysis, the mathematical algorithm that underlies Recommind's products. I am also an inventor on the patent for Predictive Coding.
4. Recommind's Predictive Coding functionality in the Axcelerate™ document review product used in the instant matter combines a proprietary, advanced text categorization algorithm with an iterative document review workflow to create probability scores that unknown documents are similar in concept to a "seed set," that is, documents known to be relevant. Predictive Coding also includes a human quality-assurance mechanism through built-in statistical sampling.
5. The purpose of this declaration is to clarify—and at times, correct—for the Court certain statements made by Plaintiffs' expert Paul Neale in the Declaration of Paul J. Neale in Support of Plaintiffs' Rule 72(a) Objection to the Magistrate Judge's February 8, 2012 Discovery Rulings (Document # 95) that I believe are or may be misleading or factually incorrect.
6. I consulted with counsel for MSL regarding the workflow set forth the ESI Protocol at issue in the instant matter, including, but not limited to, the appropriate number of iterative reviews as well as the benefit of the relevance ranking functionality of the software.

7. In ¶16 of the Neale declaration, Neale misstates the formula for recall. The numerator is currently “number of documents predicted to be responsive.” It should actually be “number of actually responsive documents identified as responsive.” The element of prediction includes false positive results, which could result in a recall calculation over 100% (an impossibility).
8. In ¶37 of the Neale declaration, he states that the formulas used for calculating recall and precision “require” an initial assessment of the number of relevant documents. While a valid approach, it is by no means the only approach, and is therefore not a requirement. Precision and recall estimates can be established after the fact by reviewing random samples from the predicted responsive set (to gauge precision) and non-responsive set (to gauge recall).
9. In ¶¶44-52 of the Neale declaration, Neale claims that the protocol fails to correctly estimate recall via a “yield-based approach” because of defects in the initial sample set in its size and makeup. A precise estimate of recall is irrelevant, however, as long as the estimation interval is within the boundary of a standard accepted by the Court.
10. In ¶58(a) of the Neale declaration, he criticizes the sampling effort by describing a scenario where the sample size fails an acceptance test that would “pass 95% of the time when our recall was 90% or higher, but fail 95% of the time when our recall was 80% or lower.” Data cited by Judge Peck suggests that recall by the current “gold standard”—keyword searches followed by human review—is significantly lower than 80%. It appears that Neale is criticizing the sample size based on an unfair standard.
11. Neale argues that predictive coding is defective because at low responsiveness rates, it becomes difficult to accurately measure recall. Neale ignores that the inverse proportion of responsiveness rate and precision of measurement of recall is intrinsic. The same difficulty would be true for any technology-assisted review, including keyword searches.
12. Therefore, I conclude that Neale is asking the wrong question: he discusses absolute recall rates, but recall should not be judged in absolute terms. It needs to be measured compared to an alternative. The use of predictive coding as outlined in the protocol will deliver better results—which I define as higher recall at lower cost—than the currently-available alternatives.
13. In Plaintiffs’ Objections, they argue that seven iterations will not be sufficient. Based on our experience, seven iterations will likely be sufficient to identify a significant number of highly relevant documents. While each review is different, we have found that seven iterations is often sufficient to identify enough documents to make educated decisions about the cost-effectiveness of further review.

I declare under the penalties of perjury that the foregoing is true and accurate to the best of my knowledge and belief.

Dated: 03/07/2012

  
\_\_\_\_\_  
Jan Puzicha